Kwaghe International Journal of Engineering and Information Technology

ISSN : 0000-0000

Index: Harvard, Boston, Sydney University, Dimensions, Lens, ResearchGet Scilit, Semantic, Google Scholar, Base etc

https://doi.org/10.58578/KIJEIT.v2i2.5342

Development of a Framework for Cybersecurity Risk Assessment in the Maritime Industry Using Machine Learning Techniques

Bartholomew Idoko & Kenneth Nwankwo
Federal Polytechnic Ohodo, Enugu, Nigeria
bartholomew.idoko@fedpod.edu.ng; kenneth.nwankwo@fedpod.edu.ng

Article Info:

Submitted:	Revised:	Accepted:	Published:
Mar 5, 2025	Mar 22, 2025	Apr 2, 2025	Apr 7, 2025

Abstract

This study assesses the level of cybersecurity risk inherent in the maritime industry in order to improve process in the sector. The maritime sector has continued to witness cyber incidents due to its importance to national economy. Also, the growing dependence of the sector on information and communications technology (ICT), as a result of increased automation, has greatly exacerbated the threats. The underlying cyber infrastructure with its expanding threat landscape and vulnerabilities have also further exacerbated the risk landscape in the sector. More so, the dearth of empirical studies in this domain is an indication of knowledge gap occasioned by non-availability of empirical data on how organizations in this sector manage cybersecurity risk. That is, how organizational operations and technological assets, individuals and processes affect the sector. Thus, the study has identified and established the cybersecurity risks specific to the maritime sector and gauged the gap based on people, process and technology elements of cybersecurity. This study uses Artificial Intelligence, machine learning model in particular to carry out the assessment. The study identified how organizations applied security controls in the sector using the metrics of people, process and technology. The risk was

analysed and graded into very high, high, moderate, low and very low from the established risk factors like threat and vulnerabilities. We used k-nearest neigbour and factorization methods for model training and risk ratings. The findings showed that the maritime sector has a high cybersecurity risk rating. This knowledge and the recommendations that followed, will help deepen the understanding of cybersecurity risks in the maritime sector as well as improve maritime process, its potential effects on service delivery, national security and economic wellbeing of the nation.

Keywords: Assessment, Cybersecurity, Maritime, Machine learning, Risk

INTRODUCTION

In this age of globalization, the introduction of new technologies in the maritime sector has created loopholes otherwise unimaginable (Rouzer, 2022). Still, this continued reliance on these technologies has exposed the maritime sector to malevolent use and abuses. Because technology readily permits attacks across state and national borders, cyber attackers have generously taken advantage of the fact that they can launch their attacks from almost anywhere and are not physically close to their targets. This allows for the fast execution of operations without endangering the lives of the perpetrators and allows them to remain anonymous (Silgado, 2018). Moreover, adversaries may employ several strategies using a range of tactics to target people, companies, vital infrastructures, and governmental organizations in the maritime sector, permitting them to maximize the threat surface (Allen, 2018). Cyber-attacks resulting from threat and vulnerabilities in the maritime sector often results in commercial and business disruptions, theft of data and information, monetary and cargo loss, damage to overall company reputation, goods and the natural environment (Sotiria, 2018). The desired outcome of identifying and assessing threats and vulnerabilities is critical for determining the potential and actual cybersecurity risks to the maritime sector. The maritime sector has continued to witness cyber incidents due to its importance to our national economy. Cyber criminals' daily attempt to undermine the confidentiality, integrity and availability of assets and resources belonging to this important national sector. However, the extent of the cybersecurity risk faced by the maritime sector is unknown, due to the lack of empirical data. Evaluating the cybersecurity risk in the sector will depend on a proper understanding and articulation of the cybersecurity triad namely process, people and technology. Consequently, this study assesses the cybersecurity

risks in the maritime sector, the potential effects on national security and further gauges the knowledge gap in this sector using k-nearest neighbour which is an effective machine learning techniques.

To properly identify and manage cyber risks, every organization must understand that technology alone does not hold the silver bullet to mitigate against cyber threats and attacks (Meserve, 2017). Unfortunately, many organizations, including the maritime sector, have dedicated huge effort in providing the technology required for cybersecurity implementation and management, while paying little or no attention to the other two important components of the cybersecurity triads, that is, people and process. A proper integration of the three triads of cybersecurity – people, process and technology, by the maritime sector would not only create a near impenetrable defense against the most sophisticated threat actors, but will also ensure that this apparent and continued fixation, by the maritime sector in providing technology solutions, at the expense of the other triads, does not continually expose it to further vulnerabilities.

In this paper, content-based assessment system and grading algorithm will be used to make good decisions and recommendations to professionals and management of the maritime sector to ensure the highest grade of cybersecurity risk assessment irrespective of the volume of the training data applied. The remaining sections of the paper is organized thus; section 2 discuss threat, vulnerabilities and risk factors that could influence the security posture of the maritime sector. Section 3 showcase the proposed methodology and the model implementation procedures, section 4 dealt with results and analysis and section 5 handles the brief conclusion and recommendation.

Cybersecurity Threats, Vulnerabilities and Risks Factors

The three key important elements of any cybersecurity incidents are the threats, vulnerabilities and risks. These concepts defines the Who? Why? When? Where? How? and the extent of every cybersecurity attack in an organization. Figure 1, shows the Cybersecurity threats, vulnerability, impact and risk factors in an organization.

Figure 1: Cybersecurity Threats, Vulnerability and Risk **(**retrieved from https://www.cybervista.net/what-every-executive-really-needs-to-know-about-cyber-part-1/)

Cybersecurity Threats

A cyber threat simply refers to the source and means of a particular type of attack (Muronga et al, 2019). Examples of cyber threat includes: Phishing, Phreaking, Social Engineering, Pharming, Insider, Spies and malware (Idoko & Bush, 2023). These threats can be initiated and executed by Hackers, Hacktivist, Terrorists, Cyber Vandal, Nation State actors, etc with a specific motive either personal gain, political motive, competition, monetary gain, etc.

A threat landscape is an assemblage of dangers within a specific domain or context that includes data on threats, risks and vulnerabilities in the system, threat actors, and patterns that have been noticed (Hiller, 2017). The cybersecurity threats is landscape ever-evolving. Increased information technology product complexity, a variety of external factors like the financial crisis, new vulnerabilities, the sophistication of available tools and attacks, the availability of resources (both human and financial), the availability of skills, networks that facilitate knowledge transfer, the expansion of illicit profit in cyberspace, etc. are the main forces behind these changes. (Nordel, 2022)

Vulnerabilities

Vulnerability is define as a weakness or flaw that an asset or set of assets has that one or more threats could exploit (IBM, 2024). Any weakness or flaw that an adversary could use to obtain unauthorized access and then destroy, steal, or damage an asset is called a vulnerability. (Idoko & Bush, 2023). Vulnerabilities may arise from, but are not restricted to, deficiencies in the physical security system, present management procedures, etc. Examples of cybersecurity vulnerabilities include Broken Authentication Management, Insecure Direct Object References, Cross Site Request Forgery (CSRF), Operating System (OS) and Lightweight Directory Addressing Protocol (LDAP) Injections, Structured Query Language (SQL), Insecure Cryptographic Storage, Insufficient Transport Layer Protection, and Invalidated Redirects and Forwards. (Idoko & Bush, 2023).

Risk

Cybersecurity risk is any threat or vulnerability resulting from the use of information technology that jeopardizes the principles of data and service confidentiality, integrity, and availability (CIA) (Kure et al. 2018). Risk is the deterioration of functional technology that will ultimately result in interruptions to company operations, failures of vital infrastructure, and physical harm to people and property. Cyber risk can result from both natural and artificial sources (Allianz, 2016). Understanding the cybersecurity risk that a company faces is critical to the effectiveness of any security design.

Risk Assessment

Risk Assessment consists of three components namely asset identification, threat assessment and vulnerability assessment.

Asset Identification

The first step in identifying and comprehending the risk that a company faces is determining which of its accessible assets require protection. This entails determining the organization's essential assets, or those that significantly affect availability, confidentiality, and integrity while supporting the goals and operations of the business (Grant and Goodyear, 2021). Once these assets have been identified, you can ascertain which ones need to be safeguarded and monitored the most, as well as which ones are most vulnerable to attacks from outside parties and even authorized insiders.

Threat Assessment

This focuses on analyzing an attacker's resources and helps develop specific security policies that need to be implemented, unlike a risk assessment which focuses more on analyzing the potential and tendency of the organizations resources to fall prey to various attacks (Kimberly and Kevin, 2019). It is important that penetration testing should be performed regularly and that it forms part of the cybersecurity function or task within the organization, in order to help it determine or assess the threat profiles of the organization, thereby developing an effective countermeasure against any form of attack (Boyes, 2023).

Vulnerability Assessment

An extensive analysis of the current and potential risks as well as vulnerabilities in the systems and networks of the company is known as a vulnerability assessment. It entails identifying, assessing, and ranking any vulnerabilities in your system before hackers take advantage of them (Kimberly and Kevin, 2019). In order to assist you in strengthening these defenses, it evaluates your current cyber security policies during this process. Establishing organizational policies linked to routine vulnerability assessments is a useful strategy to lessen the impact of a potential cyber threat. This will also give the business insights into areas where its cybersecurity efforts may be strengthened.

Cybersecurity Risk Assessment Methods

Risk assessment methodologies in maritime processes can be broadly classified into two types: qualitative and quantitative. While the quantitative assessment approach statistically analyzes risk by assigning numerical risk ratings, the qualitative assessment method prioritizes specific risks through the examination of occurrence probability (Allen, 2018).

Cybersecurity Risk Scenerio

This is a description of potential occurrences that could have a positive or negative, unclear impact on the organization's goals being achieved. The following elements make up a cybersecurity risk scenario: threat kinds; actor skill, capability, and location; assets; events; and time. In order to be more prepared for a variety of future possibilities, the organization often thinks broadly when faced with particular events and develops a range of potential outcomes (Kure et al., 2018).

Proposed Model Design and Implementation Process

In this paper, an automated risk assessment system using dataset from the federal ministry of marine and blue economy is proposed. This AI enabled system is more effective and secured when compared with the manual system as it will further improve marine processes by identifying the cybersecurity risk attributed to people, process and technology. The system is designed to efficiently determine the risk value/scale using the five NIST framework core; namely identify, protect, detect, respond and recover. The dataset on metrics evaluation of the cybersecurity risk shows a pattern of the risk value for each of the participating organizations. Three major organizations within the maritime industry in Nigeria were considered in the research, namely, the Nigerian Maritime Administration and Safety Agency (NIMASA), the Nigerian Port Authority (NPA) and the Nigerian Shippers Council. The relationship and clusters determined by the data analysis in these organisations is then embedded in the machine learning module for the development of the risk assessment system.

Explored Algorithms

There are various algorithms used for risk assessment system, prominent among which are; Support Vector Machine, Naïve Bayes, Convolutional Neural Network, Random Forest, etc. However, this paper employed the use of k-nearest neighbor and matrix factorization.

K-Nearest Neighbor

This is a classification-based algorithm that could be used to classify risk instances based on their relationship or distance in relation as observed in the proposed risk matrices of identify, protect, detect, respond and recover from the dimension of people, process and technology. This is illustrated in figure 2.

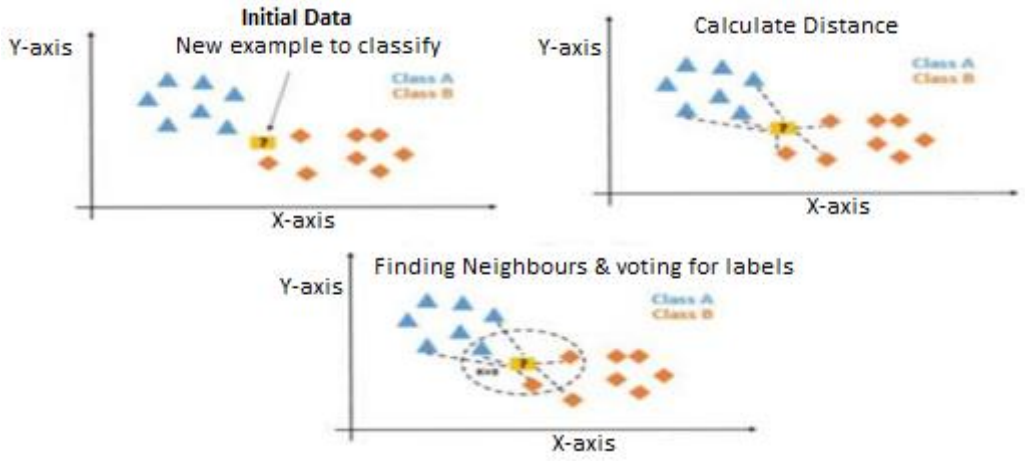

Figure 2: Structure of K-Nearest Neighbour

The distance between the class A and B of the strata is known as Euclidean Distance. This is used to predict risk based on unique features or behavior. It is used in this research for risk classification into most common types or classes. There are also risk which seems to exist on their own which are grouped and are none-linear strata of risk which will require a special random approach to their classification and recommendation algorithm. The K-Nearest Neighbor is modeled to mimic trend in the risk rating behavior of the selected agencies. There is opening on the e-marine page from which certain targeted data are collected for the machine learning purpose. This includes; Asset Management, AM - used to determine the asset value in the organization. Awareness and Training, A&T - use to determine how prepared the organizations are to face cybersecurity challenges. Analysis and Communication, A&C - used to responds to risk scenario. Anomalise and Detection, A&D – use to detect risk before an attack on the maritime infrastructures occur.

The adjectival keywords in the review body are usually targeted by the machine learning algorithm to determine the rate at which the risk is managed or mismanaged by the agencies. Keywords such as (low, very low, moderate, high, very high) were targeted in the body content of the review to determine what the agencies speculate on the risk bedeviling them. There are also many unverified risks and vulnerabilities such as Zero day vulnerabilities that also exist, contributing significantly to the overall distribution of data. While their impact is not felt by the organization because the vulnerability is unknown, the risk factor is high because the agencies will find it difficult to prevent and recover hence, it is important that we consider the risk as a result of zero day vulnerability as a valuable part of the machine learning design.

Matrix Factorization

Matrix factorization require a breakdown of actual risk classifier matrix into two, having one as the upper triangle (U) while the other as a lower triangle (L). What this implies is that the Upper (U) and Lower (L) angles of the matrix represent different classes of risk but with common general class. This is mostly used in collaborative filtering method in risk assessment algorithm.

Dataset Information and Analysis

The dataset used in this paper is made up of incidences that could potent cybersecurity risk for the maritime sector (NIMASA, NPA and NSC). The dataset comprises of 22,105 rows and 10 columns collected between the period of 12-02-2018 and 30-12-2023 with the following columns; job role, IT strategy plan, IT policy, risk assessment frequency, security training, security awareness, risk rating, review headline, review body, and review date respectively. Review_headline shows the title of the review, review_body contains the context of the review, and review_date is the date the review was written. For the purpose of this assessment, we concentrated on the curated dataset of the NIMASA, NPA, NSC review which we believe would provide a greater percentage of concrete and quantitative assessments as compared to other agencies in the maritime sector.

Feature Extraction

The dataset provides some features which are not needed for the purpose of training this model. In this regard, agency id, risk_rating, review_date are selected as they are most useful to this evaluation and the rest were cast-off. The dataset was examined for missing values, and they were found to be all complete. Furthermore, agency_id made up of alphanumeric values of object datatype was transformed to integer values to help lessen feature complications. Dataset preprocessing was performed to convert or encode data to a state that machine can deconstruct so easily for its algorithm. The essence is to normalize the data before experimenting with it.

A step took to normalizing the dataset was to transform the agency_id column of object datatype made up of alphanumeric values to a unique integer value to reduce the complexity and to avoid irregularity in the process during the classification process and date of reviewing its timestamp equivalent. Some agencies have over reactively provided high risk ratings and least risk ratings while some have rated low. To normalize these

ratings for those kinds of agencies, data of agencies who have rated 30 risks and more were considered.

Results and Analysis

In this section, we analyze the result of our developed model using simulation in order to showcase the efficacy of our k-nearest neighbour algorithm.

Simulation

We developed our model in order to train the data collected based on features embedded in maritime cyber threats, vulnerabilities and risk factors been the most contributing variables. The dataset was preprocessed and trained. Supervise and Single Value Decomposition (SVD) is used to disintegrate the original matrix and latent factor that helps in generating the assessment ratings. In order to achieve a good accuracy in this assessment, we presented a carefully selected features as part of the post-processing. Table 1, figures 2 and 3 represent data sample, risk rating output, and mean distribution of risk rating respectively.

Table 1: Data sample

S/N	Agency_id Threat_id		Vulnerability_id	risk_rating	timestamp
0	NIMAS8U	239456	132793	5	13658200
1	NPAXL7O	356811	321732	5	13411000
2	NSCOO3W	543189	439886	1	13671900

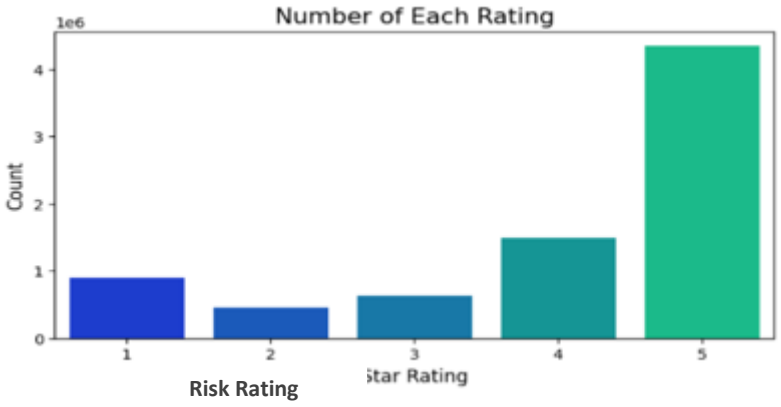

Figure 3: Risk Rating Output

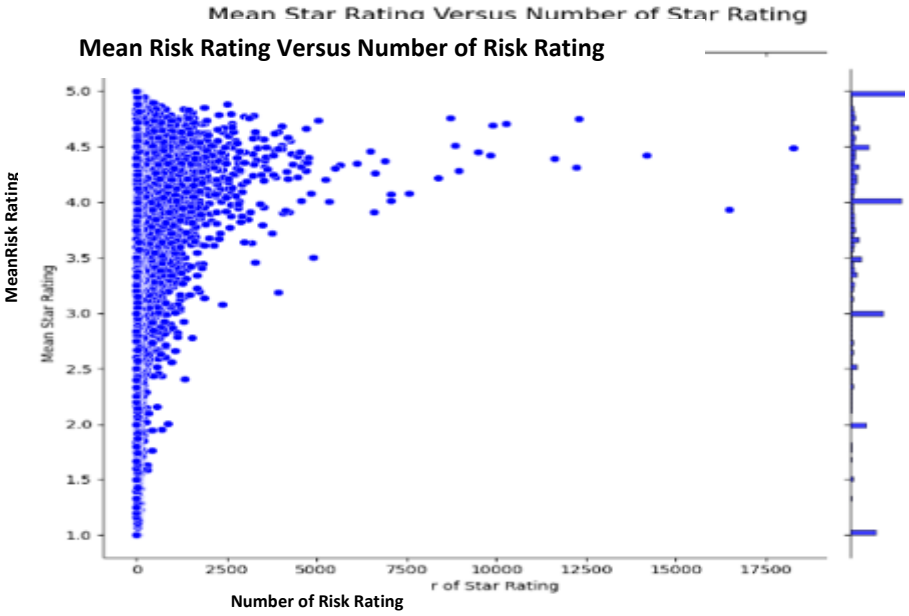

Figure 4: Mean Distribution of Risk Rating

We made our assessment using collaborative and content-based learning with agencies who have over 40 vulnerabilities. The training data from this model contains 79,105 observations, 1, 500 vulnerabilities and 31, 250 threats. Also risk rating given by an agency to a threat and vulnerability in a given period was ascertained. Using the threat-vulnerability similarity, the proposed system made top five assessments for five 3 agencies as depicted in figure 4.

agency_id	risk_id	score	rank
NIMAS8U	239456	0.044481935501098635	1
NIMAS8U	239456	0.03829501509666443	2
NIMAS8U	239456	0.034575473070144656	3
NIMAS8U	239456	0.031499905586242674	4
NIMAS8U	239456	0.027724857330322265	5
NPAXL7O	356811	0.044481935501098635	1
NPAXL7O	356811	0.03829501509666443	2
NPAXL7O	356811	0.034575473070144656	3
NPAXL7O	356811	0.031499905586242674	4
NPAXL7O	356811	0.027724857330322265	5
NSCOO3W	543189	0.044481935501098635	1
NSCOO3W	543189	0.03829501509666443	2
NSCOO3W	543189	0.034575473070144656	3
NSCOO3W	543189	0.031499905586242674	4
NSCOO3W	543189	0.027724857330322265	5

Figure 5: Risk Assessment

Performance Evaluation

In predictions, accuracy metrics are generally applied for system valuations, Root Mean Square Deviation (RMSD) also known as Root Mean Square Error (RMSE) is used to determine the distinction between the predicted values and the assessment of the actual values as shown in equation 1 (Bush, et al., 2018).

$$\text{RSME} = \sqrt{\frac{1}{n}} \sum_{i=1}^{n}(|xi - yi|)\text{RSME} = \sqrt{\frac{1}{n}} \sum_{i=1}^{n}(|xi - yi|) \quad (1)$$

The proposed system makes recommendations according to user behavior by predicting cyber risk ratings in maritime processes. The four potential outcomes of the predictions are:

- True-Positive: this indicates situations where the outcome of the model correctly predicted the risk as positive.
- False-Positive: represents situation where the model falsely predicted the risk outcome as positive.
- False-Negative: this implies in situations where the model predicted the risk outcome as negative.
- True-Negative: this indicates situations where the outcome of the model correctly predicted the risk as negative.

Precision (P) calculates the fraction of the accurate positive predictions on the total numbers of predictions thus:

$$P = \frac{True\ Positives}{True\ Positives + False\ Positives} P = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

Recall (R) is used to calculate the fraction of the positive predictions against the sum of the true positive values. The true positives here are the combination of the true positives and false negatives as shown in equation 3:

$$R = \frac{True\ Positives}{True\ Positives + False\ Negatives} R = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

This paper explored both collaborative and content-based assessment. The collaborative filtering had RMSE of 4.42 and the hybrid assessment produced RMSE and MAE of 0.32 and 1.20 approximately. Figure 5 shows the classification report of the propose system.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	9
2	1.00	0.25	0.40	8
3	0.50	1.00	0.67	6
4	1.00	1.00	1.00	42
5	1.00	1.00	1.00	136
accuracy			0.96	201
macro avg	0.90	0.85	0.81	201
weighted avg	0.99	0.96	0.96	201

Figure 6: Classification report of the proposed system

The train data shows a consistent increase in the rating of cybersecurity risk in the Nigeria maritime sector. It also shows a significant increase in the rating for some agencies at the tail end of the analysis which comes with emerging technologies in 2022.

Figure 6 shows an accuracy of 0.96 which is very significant. This means the accuracy of our risk rating in our context of people, process and technology is relative high. As shown in table 2, result of the proposed system is compared with the performance of some existing systems, and it is evident that the robustness of the proposed system resulted to its higher accuracy.

Table 2: Comparative results analysis

Sources	**Methods**	**Performances (Accuracy)**
Dooms S., et al (2018)	UserKNN	0.91
S.G.K. Patro et al (2021)	HAR-KNN	0.86
Salam P., et al (2022)	Pearson correlation similarity	0.90
Proposed system	KNN+SVD	0.96

Conclusion

The maritime sector is exposed to cybersecurity risk at high level, as a result of the inherent vulnerabilities and threats in the system and these have been left undetected. Most attack on the maritime sector is usually directed first at the vulnerabilities in the organization system. These vulnerabilities and threats are possible because of the failures in the cybersecurity triad, namely people, process and technology. Unfortunately, these vulnerabilities, threats and by extension the cybersecurity risk in the maritime sector is not abetting, but rather growing worse. The maritime sector must therefore be proactive in ensuring that every effort is made to further reduce its threat surface and cybersecurity risk

exposure, to ensure that the confidentiality, integrity and availability of systems, networks and data is preserved and protected.

The recommendations from this assessment when implemented, will assist the maritime sector in closing the knowledge gap regarding cybersecurity risk present in the sector by providing empirical facts to deepen the understanding of cybersecurity risks in the maritime sector, its potential effects on service delivery, national security and economic wellbeing of the nation.

We recommend access to critical data be made available for research in the sector in order to effectively improve the existing cybersecurity governance framework and structure within the maritime sector so as to ensure that an informed decision-making process is achieved and to assist the organizations in identifying and managing cybersecurity risks.

Acknowledgment

The authors would like to express our sincere gratitude and appreciation to the Tertiary Education Trust Fund (TETFUND) Nigeria, for providing the funding for the research.

References

Allen C. H. (2018). Developing and Implementing a Maritime Cybersecurity Risk Assessment Model. Retrieved 23 July, 2024, from: https://ssrn.com/abstract=3302772

Allianz, G. (2016). Allianz Risk Barometer: Top Business Risks 2016. Munich: Allianz SE and Allianz Global Corporate & Specialty SE.

Boyes, H. (2013): Maritime Cyber Security – Securing the Digital Seaways.

Bush, I., Abiyev, R., Ma'aitah, M. and Altıparmak, H. (2018). "Integrated artificial intelligence algorithm for skin Detection". In ITM Web of conferences EDP Sciences Vol. 16, pp. 02004.

Grant, S. & Goodyear, J. (2021) ECDIS: Past, Present and Future. Science Review 2020 and 2021, Department of Fisheries and Oceans, Dartmouth.

Hill, M. (2020): Danish Railway Company DSB Suffers DDoS Attack. Retrieved 20 July, 2024, from https://www.infosecurity-magazine.com/news/danish-railway-ddos-attack/.

Hiller, A. (2017): The Challenge of Cybersecurity in the Maritime Domain

IBM. (2024).Zero day Attack. Retrieved 23 July, 2024, from https://www.ibm.com/topics/zero-day

Idoko, B. & Bush, I. (2023). IoT Security Based Vulnerability Assessment of E-learning

Systems. Machine Learning and the Internet of Things in Education, Studies in Computational Intelligence 1115, https://doi.org/10.1007/978-3-031-42924-8_15. Springer e-book.

Kimberly T. & Kevin D. J. (2019). A Model-Based Framework for Maritime Cyber- Risk Assessment, WMU Journal of Maritime Affairs, DOI: 10.1007/s13437-019-00162-2.

Meserve, J (2017) Sources: Staged Cyber Attack Reveals Vulnerability in Power Grid. CNN. Retrieved 05 June, 2024 from www.cnn.com/2007/US/09/26/power.at.risk.

Muronga K; Letebele M, Binda P, and Smith-Godfrey S. (2019) - Towards Secure Maritime Transport in South Africa: an investigation of cybersecurity readiness of organizations, available at CSIR Built Environment, Transport Systems and Operations.

Nordell, D. (2022). Cyber and Technology Threats to the Tanker Industry. London: CSCSS (Centre for Strategic Cyberspace and Security Science).

Rouzer, B. (2022). Evolving Cyber Threats to the Marine Transportation System. Washington: US Coast Guard.

Silgado D. M. (2018). Cyber-attacks: a digital threat reality affecting the maritime industry.

Sotiria Lagouvardou, (2018). Maritime Cyber Security: Concepts, Problems and Models. Vol. 2. London Press, TPG, 1674, PP. 24-53.